SCLD: a stem cell lineage database for the annotation of cell types and developmental lineages

Edward E. Hemphill[1], Asav P. Dharia[1], Chih Lee[2], Caroline M. Jakuba[1], Jason D. Gibson[1], Frederick W. Kolling IV[1] and Craig E. Nelson[1],*

[1]Genetics & Genomics Program, Department of Molecular & Cell Biology, University of Connecticut, Unit 2131, Storrs, CT 06269-2131 and [2]Department of Computer Science & Engineering, University of Connecticut, Unit 2155, Storrs, CT 06269-2155, USA

ABSTRACT

Stem cell biology has experienced explosive growth over the past decade as researchers attempt to generate therapeutically relevant cell types in the laboratory. Recapitulation of endogenous developmental trajectories is a dominant paradigm in the design of directed differentiation protocols, and attempts to guide stem cell differentiation are often based explicitly on knowledge of *in vivo* development. Therefore, when designing protocols, stem cell biologists rely heavily upon information including (i) cell type-specific gene expression profiles, (ii) anatomical and developmental relationships between cells and tissues and (iii) signals important for progression from progenitors to target cell types. Here, we present the Stem Cell Lineage Database (SCLD) (http://scld.mcb.uconn .edu) that aims to unify this information into a single resource where users can easily store and access information about cell type gene expression, cell lineage maps and stem cell differentiation protocols for both human and mouse stem cells and endogenous developmental lineages. By establishing the SCLD, we provide scientists with a centralized location to organize access and share data, dispute and resolve contentious relationships between cell types and within lineages, uncover discriminating cell type marker panels and design directed differentiation protocols.

INTRODUCTION

The field of stem cell biology and the associated data has vastly expanded as more and more researchers attempt to generate therapeutically relevant cell types in the laboratory. Because most attempts to guide stem cell differentiation are based on *in vivo* development (1), stem cell biologists rely heavily upon developmental biology data such as cell type-specific gene expression profiles, anatomical and developmental relationships between cells and tissues and signals important for development and differentiation of stem cells to mature cell types. This data is currently accessible through a variety of resources (Publications, Databases, etc) and formats. The SCLD aims to consolidate these three key types of information into a database providing a central resource to the stem cell community.

Years of experimental embryology has generated large amounts of tissue-specific and cell type-specific gene expression data (2) that researchers can use to study cell types outside of their normal anatomical context. Cell type-specific gene expression profiles allow researchers to establish the identity of cells, provide insight to the developmental pathways activated in differentiating cells and monitor the progression of directed differentiation protocols. A wide variety of databases store this data and make it available to the entire research community. None, however, has the ability to compile data across existing databases and allows users to manipulate and edit the information.

During development, embryonic and somatic stem cells progress through a series of transitional cell types ultimately resulting in terminally differentiated cells. This developmental path is often graphically represented as a cell lineage map. Stem cell biologists rely heavily on these maps as guides when designing directed differentiation protocols to mimic endogenous development *in vitro*. Several resources are dedicated to describing developmental relationships in a variety of model species to assist in this navigation, yet none provide a complete set of cell lineages in a format that can be readily explored and exploited by a stem cell scientist. Stem cell biologists are currently forced to amass cell lineage information piecemeal from a disparate set of sources ranging from textbooks to primary scientific literature.

*To whom correspondence should be addressed. Tel: 860-486-5617; Fax: 860-486-4331; Email: craig.nelson@uconn.edu

Culture conditions and environmental cues used to direct progenitor cells toward desired lineages or terminally differentiated cell types are among the most crucial pieces of information for the successful differentiation of stem cells toward useful cell types. While an enormous amount of effort has been invested in identifying these cues *in vivo* and *in vitro*, to the best of our knowledge no existing database curates this information in a computationally accessible manner. With a majority of stem cell culture protocols being founded upon our knowledge of developmental biology, ready access to this information would be extremely valuable to the stem cell community.

A large number of databases are dedicated to supporting and disseminating gene expression data [for review, see de Boer *et al.* (2)]. Some of the expression databases useful to a stem cell biologist include GXD (3) from Mouse Genome Informatics (MGI) at Jackson labs (4), which annotates gene expression profiles across all of mouse development. Another example is the Novartis Gene Expression Atlas (5), which stores human tissue-specific gene expression profiles. As useful as these resources are, they are designed to benefit a wide cross-section of scientists and, for this reason, cannot optimally address the specific needs of a stem cell scientist.

One significant effort to consolidate stem cell data from across labs is the Canadian Stem Cell Network's (6) StemBase (7), which curates, organizes and stores gene expression data from genome-wide transcript profiling experiments on mouse and human stem cells and their derivatives. While StemBase provides search and analysis tools, it does not provide the ability for users to edit the data or to curate cell types and lineages.

Other recent data sets and databases do address a subset of these needs within specific cell lineages. The most well developed examples can be found for the hematopoietic lineage and include the HaemAtlas (8); the Erythropoiesis database (EpoDB) (9); LymphTF-DB, a database of transcription factors involved in lymphocyte development (10); Hematopoietic Fingerprints (11) and BloodExpress (12), which unite information on the entire hematopoietic lineage. Many of these databases combine lineage information with expression data in ways that would be advantageous to stem cell biologists if extended to additional lineages including musculoskeletal, pancreatic, neuronal, endodermal and cardiac as well as many others.

Well-defined lineage maps are required to unite gene expression data with defined stages of differentiation. Several resources describing developmental and lineage relationships in the embryo are available including the Edinburgh Mouse Atlas Project (EMAP) (13), and the Expressed Sequence Annotation for Humans (eVOC) (14), among others. Typically, these projects rely upon standardized ontologies that allow for the precise description of cell types and structures in the developing embryo. One issue with the use of these ontologies for stem cell biology is an extensive reliance upon anatomical and stage-specific relationships, with a relatively minor representation of the lineage relationships most important for the design of directed differentiation protocols. Recognition of this fact led to the initiation of Cell Type ontology (CL) (15) by Bard and colleagues, which lays the groundwork for curating lineage relationships throughout the vertebrate embryo.

Thus, no existing database brings together the data most useful to stem cell biologists in an expandable, easily accessed and annotated, user-editable interface. To address this need, we present a new resource called the Stem Cell Lineage Database (SCLD). The SCLD provides a user interface for the curation and browsing of cell type gene expression profiles, cell lineage maps and notes on the manipulations and culture conditions necessary to transition between cell types within a lineage. Our database stores information for both human and mouse cell types and lineages, is publicly accessible, directly user editable, and hyperlinked to a series of source databases, thus benefiting from the vast amounts of information presently available and maintaining consistency with current scientific nomenclature. The SCLD provides graphical lineage maps, graphical ontology building, cell type expression data and a variety of other data essential to stem cell biologists. By providing a user-editable interface for the exchange of directed differentiation protocols and sharing of stem cell research advances, the SCLD will help address the need for swift communication of new advances in the rapidly progressing field of stem cell science.

DESIGN OVERVIEW

The SCLD is an internet application accessible through internet browsers (Firefox, Internet Explorer and Safari) communicating to a computer server sending and retrieving data. The server contains both the application and database servers. The application server (Apache Tomcat) is the central hub that runs the SCLD. This provides web pages to users as well as controls the data being sent and retrieved to both the user and the database. The database (MySQL) is where the actual data for the SCLD is stored and maintained. Direct access to the database is not provided; instead it is accessed indirectly through the application server. Open source technologies have been used in the design, development and deployment of the SCLD. Additional information on the computational and database design can be found in the Supplementary Data.

Stem cell research is progressing rapidly and laboratories are often working on different cell lineages making it unlikely that a centrally maintained database would be as useful or timely as an open-source counterpart. For this reason, we have designed the SCLD as a user-editable database. Users submit a request for a user account providing them the means to add/modify the data within the SCLD. Through the user accounts, users can enter and edit the cell types, markers and lineages most relevant to their own work. Researchers can also comment on cell types and lineage maps to facilitate discussion of contentious issues. This structure avoids the strict anticipation of the specific needs of independent users; each can shape the SCLD to be most beneficial to their research goals. To maintain data integrity, any data entered manually is required to have supporting evidence from a

published reference(s) retrieved automatically from PubMed. In addition a history is maintained of all edits made to cell types and lineages. If necessary, previous edits can be restored for cell types and lineages.

Using existing biological resources, the SCLD database is populated with information such as gene names, cell type names and gene expression data. This significantly reduces the barrier to the entry for the annotation of cell types and cell lineages, prevents errors associated with manual entry and provides immediate value-added information to new users. For example, gene names for both mouse and human are automatically downloaded from MGI and the Human Gene Organization (HUGO) (16), and are periodically updated to reflect the latest changes ensuring consistency among entries.

While there are many instances where the data on cell types and lineages for mouse is similar to humans, this is not always the case. Keeping the data separated between mouse and human ensures that an accurate picture is

presented for cell types and lineages for each respective species.

CELL TYPES

Cell types are commonly identified by name and by the expression of a unique set of genetic markers (i.e. insulin is expressed solely in pancreatic β-cells). The SCLD compiles gene expression data for cell types and allows searching for cell types by name or combination of gene markers expressed or not expressed. This search function allows researchers to efficiently identify the expression profiles of known cell types, determine which genes may be used to identify a cell type, and identify candidate cell types matching known expression profiles. When searching by name, the cell type name is either typed in or selected from the list of stored cell types (Figure 1A). Alternatively, gene markers known to be expressed or not expressed can be entered and the SCLD will produce a list of candidate cell
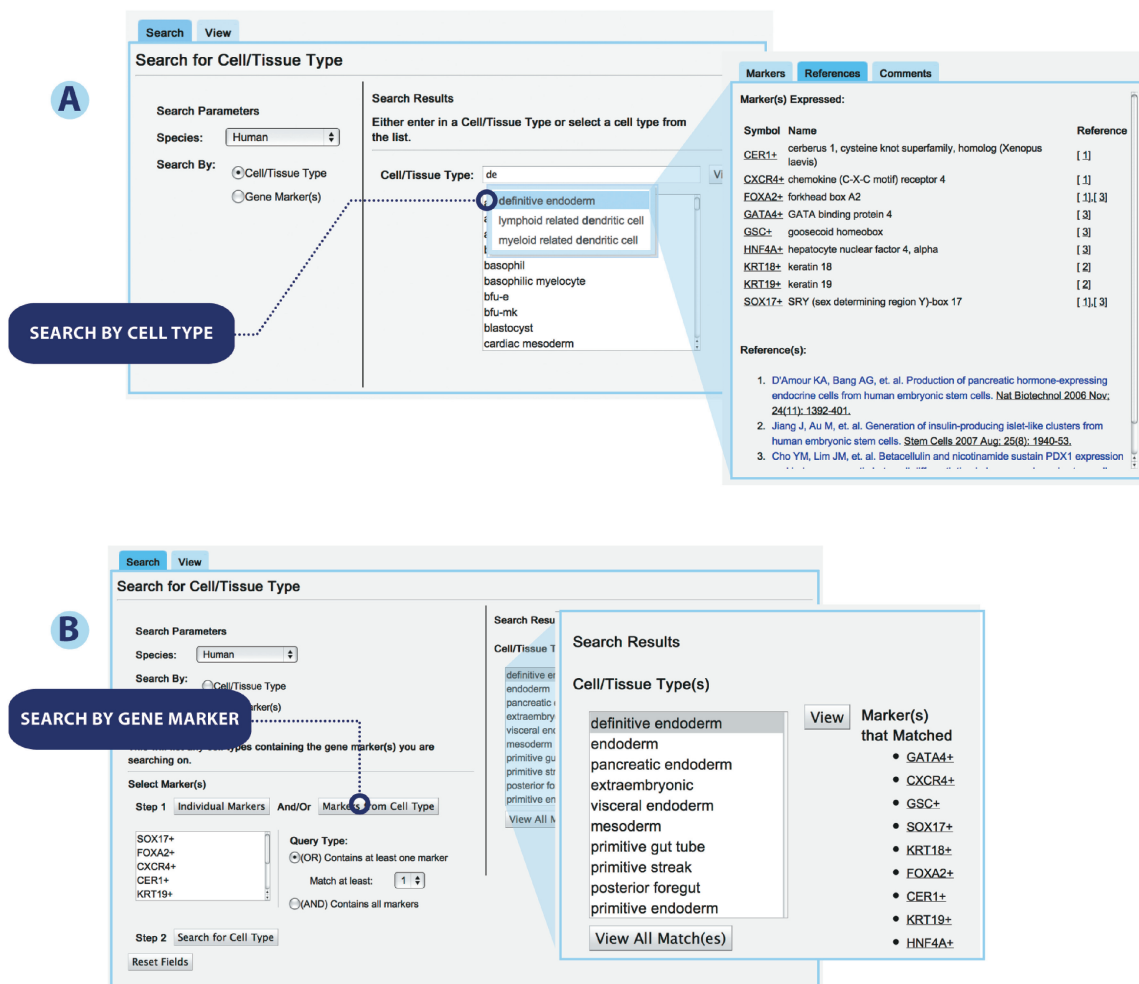
Figure 1. Searching for cell types. (**A**) Search by cell type. A cell type (i.e. definitive endoderm) is selected from a list box or can be autofilled as a user enters text. Upon selecting cell type, a list of associated gene markers and the corresponding references (right side panel) will appear. (**B**) Search by gene marker. Gene markers (i.e. SOX17, FOXA2, etc.) are queried individually or by choosing markers linked to a cell type. Searches can be narrowed by specifying all gene markers to match (AND option), or setting a minimum threshold on the number of matches (OR option). The resulting search displays the cell types along with the matching gene markers (right side panel B).The cell types are ordered by descending number of gene marker matches.

types (Figure 1B). From either of these two searches, additional information on the cell type can be viewed including: (i) species; (ii) name and description of the cell type; (iii) gene markers used to identify the cell type; and (iv) associated publications supporting the data. Each gene marker and its corresponding reference is directly linked to MGI, HGNC or PubMed for easy access to additional information.

Along with viewing previously defined cell types [such as those in EMAP (13)], the SCLD supports editing cell types and adding new, user-defined cell types. Editing a cell type allows descriptions to be updated or gene markers and references to be added or removed. The SCLD also facilitates discussion by providing space for users to comment on all data corresponding to a cell type. Using these tools, the definition of a cell type can be modified to reflect the most current available data.

In addition to leveraging existing databases and ontologies, we have extracted condensed cell type definitions from these ontologies that are more applicable to stem cell scientists. Because MGI is annotated as an anatomical lineage, cell types imported from Jackson labs are organized by developmental time point [Theiler stage (TS)]. This organization results in developmentally

(lineally) equivalent cell types represented as multiple unique entries, one for each pertinent TS, confounding the study of a single cell type across developmental time. Using the SCLD, we combined these cell types across stages into merged cell groups in a map dubbed 'Stage All'. This reduced the 16 000 entries from MGI into fewer than 5000 Stage All cell types (Figures 2A and B). Users can search Stage All cell types, EMAP cell types or user-defined cell types as desired (Figure 2C). In order to preserve interoperability, mapping relationships between cell types in discrete ontologies are maintained throughout and stored in the SCLD.

EMAP also separates lineally equivalent cell types across serially homologous anatomical boundaries within the embryo (such as somites, branchial arches and fore and hind limbs). In order to group these cell types into merged cell types more representative of the lineal relationships of interest to a stem cell biologist, we developed a Python GUI called Cell Type Condenser (CTC). CTC makes use of the 'part of' relationships in EMAP and displays a cell type's path from itself to the root, allowing users to easily identify developmentally equivalent cell types and condense them into a single new cell type within the database. This mapping from a
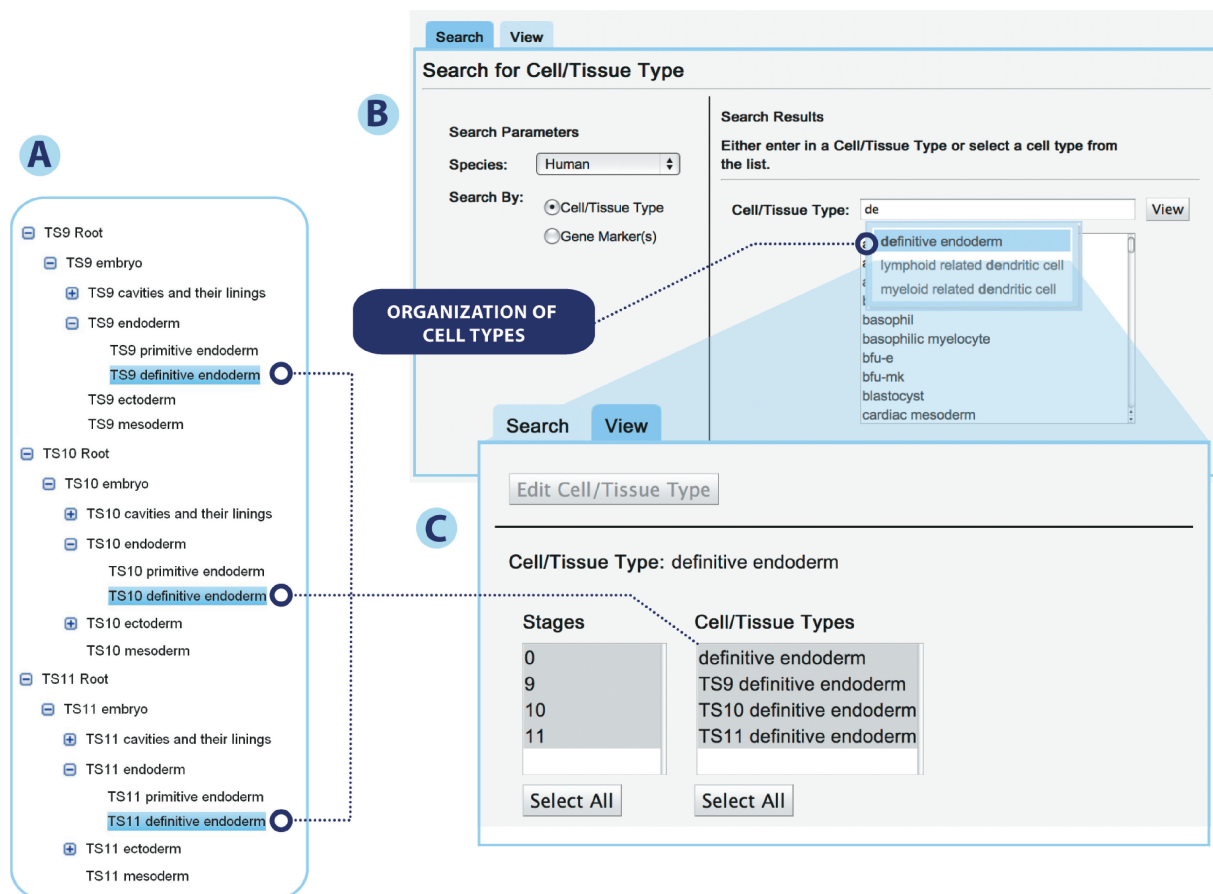
Figure 2. Stage all cell types. (**A**) Three Theiler Stages 9–11 are shown for definitive endoderm. Cell types having the same names and anatomical hierarchies within each stage are merged. Definitive endoderm defines a single 'Stage All' cell type as it has the same name and anatomical hierarchy in all Theiler stages. (**B**) The cell type (i.e. definitive endoderm) is initially presented to the user as a stage all cell type and can be expanded to reveal all the constituent cell types as shown in (**C**).

single cell type to a group of cell types enables existing and new annotations to propagate from one ontology or lineage to the other. CTC is available at the SCLD website, and saved files containing mapping information can be uploaded to and processed by the SCLD.
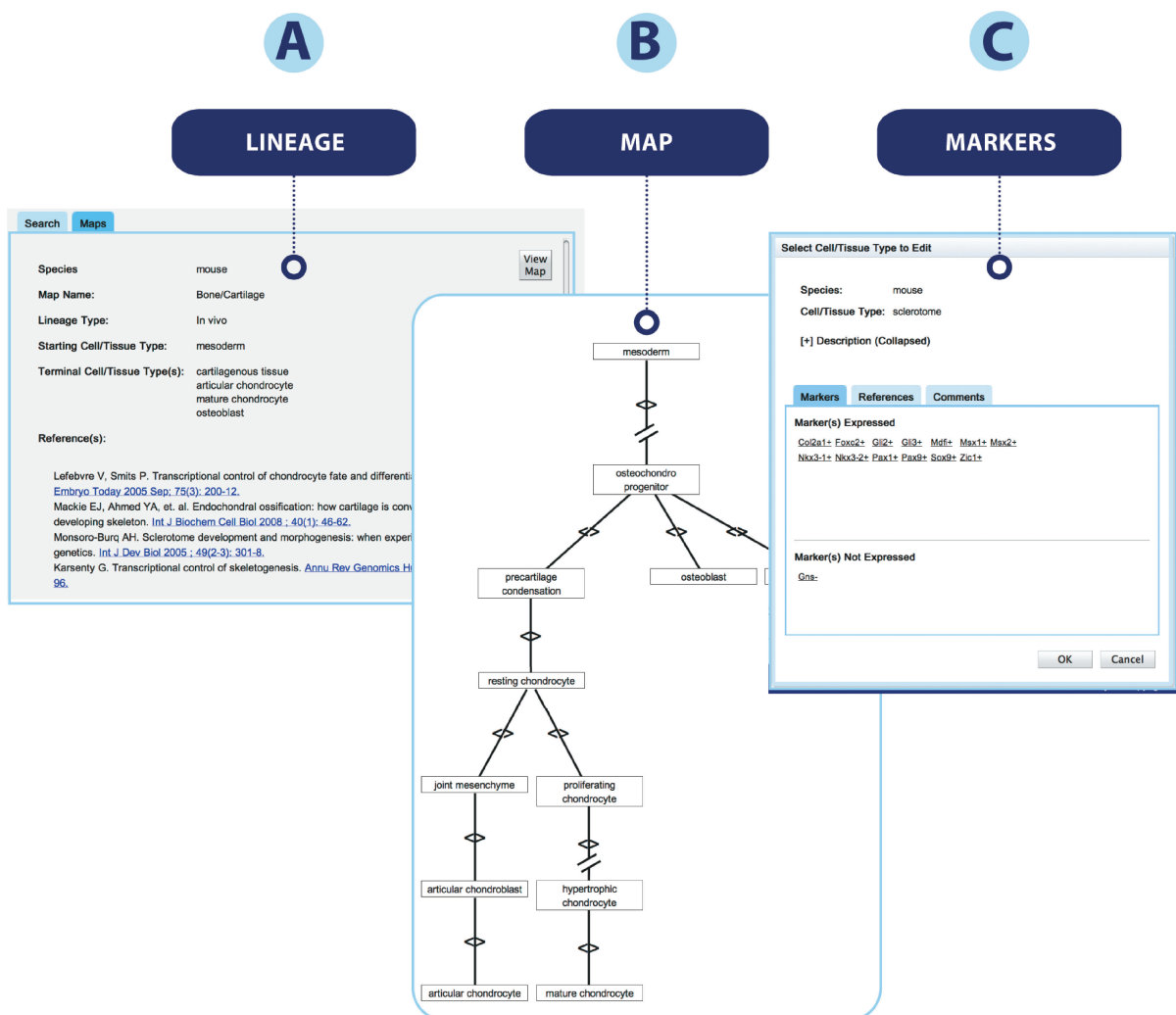
LINEAGE MAPS

Cell lineage maps represent the other major component of the SCLD. Two types of lineage maps are supported: *in vivo* and *in vitro* (see Figures 3B and 4A for examples). These maps represent lineal relationships between individual cell types that allow the user to reorder cell types from combined anatomical/developmental ontologies like EMAP into exclusively lineage-based relationships. Stored lineages are defined by the following characteristics: (i) map type (*in vivo* or *in vitro*); (ii) description, including

starting and terminal cell types; and (iii) published references supporting the data (Figure 3A).

To find, view or edit a lineage map, the SCLD supports searching based off the type of map, *in vivo* or *in vitro*, and/or any cell types contained in a map. For *in vivo* lineages, the SCLD can display the lineage map in a vertical (Figure 3B) or horizontal (Figure 4A) graphical format showing the cell types and their developmental trajectory from progenitor (top or left) to terminal (bottom or right). Highlighting a cell type on a lineage displays information on the gene markers and associated references for that cell type stored in the database. Gene names and references are hyperlinked to MGI, HGNC and PubMed (Figure 3C).

As such, the SCLD lineage interface serves as a powerful and intuitive graphical ontology editor where users can access cell types mined from other databases and align them into a developmental lineage complete

Figure 3. Lineage map. Two types of lineage maps are supported. In this example, an *in vivo* bone/cartilage lineage from our case study is shown. (**A**) Information about the lineage map is displayed including the species, type of lineage map, starting and terminal cell types, along with supporting references used to build the lineage. (**B**) Lineages are displayed as graphical maps showing the progression from a starting cell type to the terminal cell types. For example, the bone/cartilage lineage begins with mesoderm and displays developmental pathways leading to its terminal cell types, cartilagenous tissue, osteoblast, articular chondrocyte and mature chondrocyte. The symbol "//" denotes a collapsed portion of the lineage map. (**C**) Selecting a cell type (i.e. sclerotome) within a lineage map displays additional information including gene markers and their associated references.

with gene expression profiles, transition properties, experimental manipulations required for directed differentiation and user notes.

TRANSITION PROPERTIES

Specific environmental cues can cause cells to divide or differentiate, or to adopt one fate over another, and this knowledge is essential for effectively guiding stem cells toward therapeutically relevant cell types for use in regenerative medicine. The relationship between cell types in a lineage includes the established 'derives from' or 'develops from' relationships, but the SCLD has the added value of including information on the amount of time, reagents, physical manipulations and other experimental criteria required for transition between cell types. As shown in the *in vitro* lineage map example (Figure 4), selecting a transition symbol (<>), will display the conditions used to generate the next cell type. In order to facilitate debate and discussion, any of the transition properties between cell types can be modified, added or commented on by additional users. To the best of our knowledge, *in vitro* lineage maps including transition properties are not curated in any other database. Ready access to this data will facilitate the comparison and optimization of difficult steps in directed differentiation protocols.

DATA

To date, the SCLD is comprised of more than 50 000 gene markers and 5000 cell types supported by nearly 4400 published references (Table 1). This database was rapidly populated by utilizing gene expression profiles catalogued within existing data resources that are extensive for mouse, but less comprehensive for human cell types.

CASE STUDY: ANNOTATING MESODERM DEVELOPMENT

To test the functionality of the SCLD, our laboratory curated *in vivo* lineage maps for the primary mesoderm-derived cell lineages outlined in Figure 5.

Table 1. Database contents

	Human	Mouse
Gene marks	19 373	34 542
Cell types	98	5008
Lineage maps	10	19
Refences	27	4458

Breakdown of the current data contained in the SCLD.
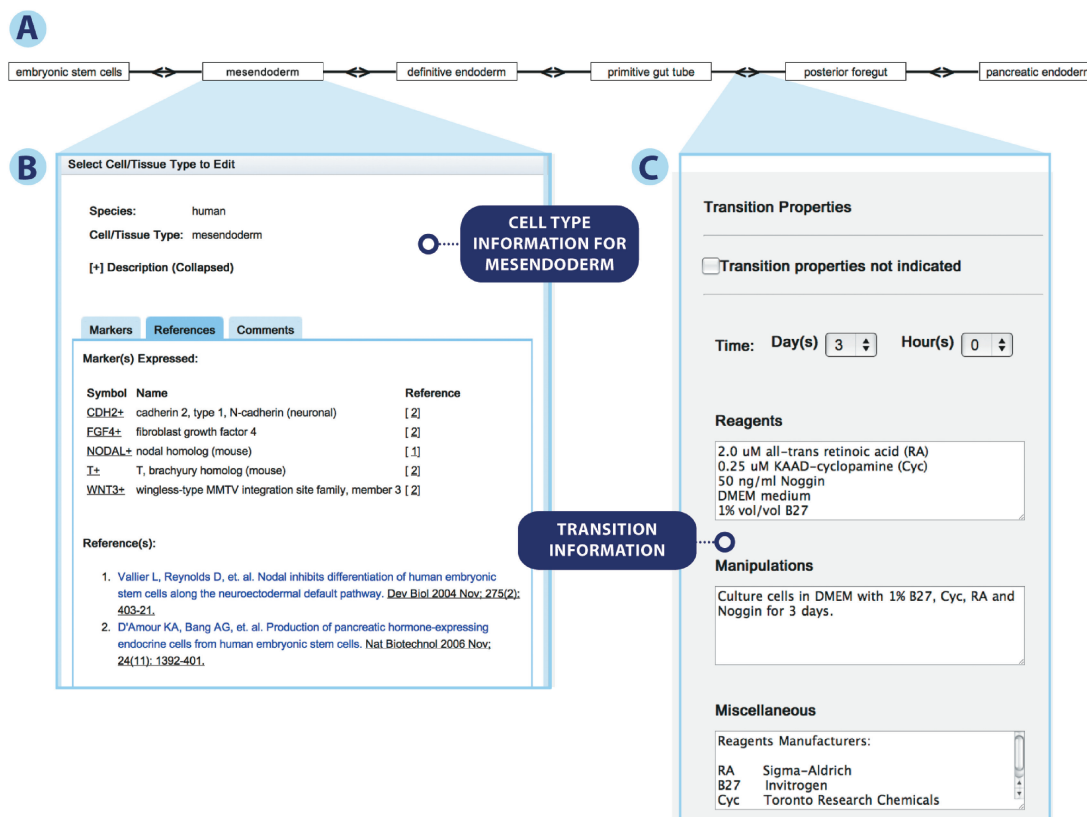
Figure 4. Interactive *in vitro* lineage map. (**A**) An *in vitro* lineage map showing the steps involved in deriving pancreatic endoderm cells from human embryonic stem cells. (**B**) Selecting a cell type (i.e. mesendoderm) in panel A displays the associated gene markers presented in panel B. (**C**) Along with the cell type information, the transition properties involved in the development from one cell type to another (i.e. primitive gut tube to posterior foregut) can be displayed by selecting the transition symbol '<>'. Information displayed includes the names of transitioning cell types, the amount of time required, reagents required, any manipulations needed as well as any miscellaneous notes.
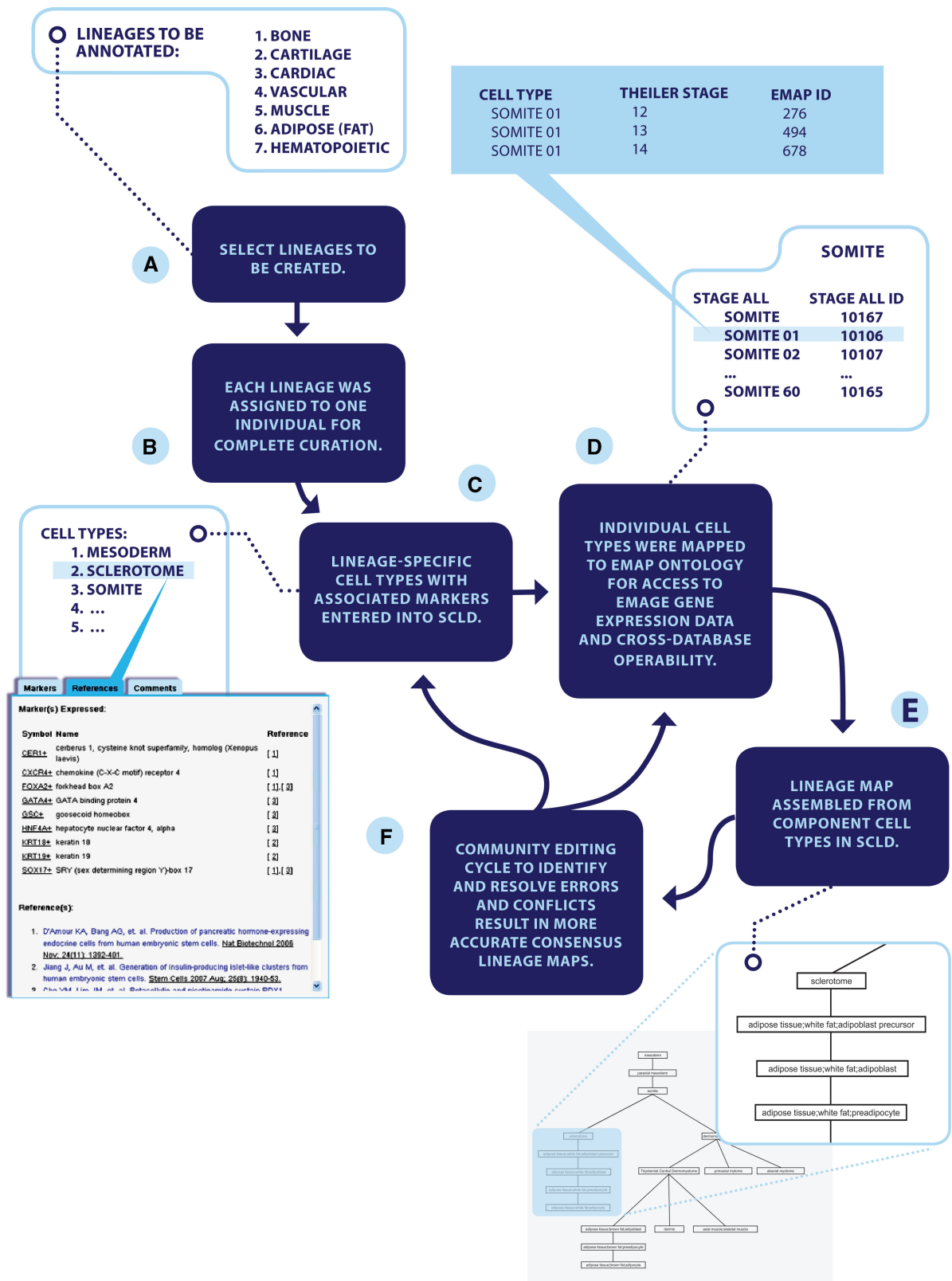
Figure 5. Case study: annotating *in vivo* mouse mesoderm development. (**A**) Choose lineages to be annotated. (**B**) Lineages were assigned to individuals for curation. (**C**) In each lineage, relevant cell types were identified and the corresponding gene markers entered into the SCLD based on published references. (**D**) Existing EMAP cell types are mapped to the individually annotated cell types (Stage All) for cross-database interoperability. The Stage All cell types are further collapsed into group cell types for easier access to gene expression data and to further reduce redundancy. (**E**) The next step involved the assembly of each cell type into an ordered lineage map. (**F**) The final step involves the SCLD users providing comments and adding to all the cell types and lineages to create a collaborative environment and to provide the most accurate lineage maps to the research community.

Each of the seven mesodermal sub-lineages (bone, cartilage, cardiac, vascular, muscle, adipose and hematopoietic) was assigned to one lab member for curation. Referencing the primary literature, the curator established the critical cell types for their sub-lineage and updated existing relevant cell types or created new cell types with names, gene markers and references in the SCLD. Curators used Stage All cell types and the CTC to link individual EMAP cell types to specific cell types in each lineage. This allowed the SCLD to populate each cell type with known gene expression data from GXD. When available, additional markers were added from relevant publications. Then, using the lineage map editing function in the SCLD, the curator arranged the relevant cell types into a lineage map representing the major developmental transitions and decisions that occur from the mesodermal progenitor to the terminally differentiated cells in each sub-lineage.

In total, this mapping of mesoderm development covered seven sub-lineages, 112 cell types and 242 gene expression markers, thus summarizing a significant percentage of known mesoderm cell types and developmental lineages. By establishing maps of these mesodermal sub-lineages, SCLD users can now visually explore the relationships between the major mesodermal lineages and cell types, readily access and download gene expression profiles for mesodermal progenitors and differentiated cells, easily map the mesodermal sub-lineage to major developmental ontologies, design directed differentiation protocols based upon these *in vivo* lineage maps, and debate and discuss the finer points of the annotated developmental relationships with other SCLD users around the world.

This case study took approximately 5 h of data assembly, cell type curation, ontology mapping and lineage building by each of seven individuals. It should be noted that none of these individuals was an expert on the chosen lineages, and that this time investment is likely to be even smaller for experts, especially when annotating the endogenous developmental cues that comprise the *in vivo* transition properties (not included in the described case study).

The very rapid manner in which data can be entered and stored for use is an indication of the low barrier to entry and the potential for true community shaping of this stem cell resource. With the support of the general stem cell community, primary literature will be quickly catalogued in the SCLD to produce a unifying reference source for all stem cell biologists.

DISCUSSION

Over the past decade, stem cell biology has become one of the most rapidly advancing fields in biology generating massive amounts of experimental data. A variety of databases have been established to organize, store and distribute this data, yet no resource has been developed that brings together the key components of developmental biology most helpful to a stem cell biologist. The user-friendly SCLD provides the stem cell field with a customizable framework to easily edit and navigate gene expression data, lineal relationships and transition properties between cell types. Its power lies in the ability to rapidly and effectively construct developmental lineages relevant to the user and serve as the basis for hypothesis building and experimental design as demonstrated by our case study.

FUTURE DIRECTIONS

We will continue to expand the SCLD's compatibility so that we can utilize future cell type ontologies, genome-wide expression databases and literature sources and maintain the most current data in the database. Another important goal will be programming the automatic extraction of unique marker profiles for specific cell types with estimates of discriminative power to aid in *in vitro* cell type identification. Along these lines, we plan to provide an API providing direct read-only access to the database. This will provide better integration and interoperability among current biological databases as well as providing an interface allowing users to write bioinformatic programs to analyze the data according to their own needs. We also intend to implement private lab-specific domains that will allow users to manipulate their data in the SCLD and store proprietary data until publication. Finally, we expect to expand the user communication and annotation functions of the SCLD in order to facilitate discussion between scientists researching complementary problems. This will include automatic flagging of related lineages and apparent conflicts between related cell types and lineages to exploit the power of social networking between investigators for resolving persistent areas of uncertainty.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Murry,C.E. and Keller,G. (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell*, **132**, 661–680.

2. de Boer,B.A., Ruijter,J.M., Voorbraak,F.P. and Moorman,A.F. (2009) More than a decade of developmental gene expression atlases: where are we now? *Nucleic Acids Res.*, **37**, 7349–7359.

3. Smith,C.M., Finger,J.H., Hayamizu,T.F., McCright,I.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Ringwald,M. (2007) The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res.*, **35**, D618–D623.

4. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Blake,J.A. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.

5. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

6. Willemse,L., Lyall,D. and Rudnicki,M. (2008) Catalyzing stem cell research. *Regen. Med.*, **3**, 761–764.

7. Porter,C.J., Palidwor,G.A., Sandie,R., Krzyzanowski,P.M., Muro,E.M., Perez-Iratxeta,C. and Andrade-Navarro,M.A. (2007) StemBase: a resource for the analysis of stem cell gene expression data. *Method Mol. Biol.*, **407**, 137–148.

8. Watkins,N.A., Gusnanto,A., de Bono,B., De,S., Miranda-Saavedra,D., Hardie,D.L., Angenent,W.G., Attwood,A.P., Ellis,P.D., Erber,W. *et al.* (2009) A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood*, **113**, e1–e9.

9. Stoeckert,C.J. Jr, Salas,F., Brunk,B. and Overton,G.C. (1999) EpoDB: a prototype database for the analysis of genes expressed during vertebrate erythropoiesis. *Nucleic Acids Res.*, **27**, 200–203.

10. Childress,P.J., Fletcher,R.L. and Perumal,N.B. (2007) LymphTF-DB: a database of transcription factors involved in lymphocyte development. *Genes Immun.*, **8**, 360–365.

11. Chambers,S.M., Boles,N.C., Lin,K.Y., Tierney,M.P., Bowman,T.V., Bradfute,S.B., Chen,A.J., Merchant,A.A., Sirin,O., Weksberg,D.C. *et al.* (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*, **1**, 578–591.

12. Miranda-Saavedra,D., De,S., Trotter,M.W., Teichmann,S.A. and Gottgens,B. (2009) BloodExpress: a database of gene expression in mouse haematopoiesis. *Nucleic Acids Res.*, **37**, D873–D879.

13. Baldock,R.A., Bard,J.B., Burger,A., Burton,N., Christiansen,J., Feng,G., Hill,B., Houghton,D., Kaufman,M., Rao,J. *et al.* (2003) EMAP and EMAGE: a framework for understanding spatially organized data. *Neuroinformatics*, **1**, 309–325.

14. Kelso,J., Visagie,J., Theiler,G., Christoffels,A., Bardien,S., Smedley,D., Otgaar,D., Greyling,G., Jongeneel,C.V., McCarthy,M.I. *et al.* (2003) eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.*, **13**, 1222–1230.

15. Bard,J., Rhee,S.Y. and Ashburner,M. (2005) An ontology for cell types. *Genome Biol.*, **6**, R21.

16. Eyre,T.A., Ducluzeau,F., Sneddon,T.P., Povey,S., Bruford,E.A. and Lush,M.J. (2006) The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res.*, **34**, D319–D321.